2024-2 시계열자료분석팀 클린업 2주차

[목차]

I. 모형의 식별

1. 시계열 모형의 필요성

본격적으로 시계열 모형에 대해 배우기에 앞서, 그 필요성에 대해 알아보도록 하겠습니다. 오차항 Y_t의 공분산 행렬은 다음과 같습니다.

$$\Gamma = \begin{pmatrix} Cov(Y_1,Y_1) & Cov(Y_1,Y_2) & ... & Cov(Y_1,Y_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(Y_n,Y_1) & Cov(Y_n,Y_2) & \dots & Cov(Y_n,Y_n) \end{pmatrix}$$
$$= \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \dots & \gamma(0) \end{pmatrix}$$

만약 Y_t가 백색잡음이라면, 대각요소를 제외한 요소들은 모두 0이 되어 분산만 추정해주면 되지만, 백색잡음 과정이 아니라면 그 이외의 요소를 모두 추정해주어야 합니다. 이때 앞으로 배울 '모형'을 활용하여 추정을 하게 되는 것입니다.

즉, 회귀/평활/차분 등의 방법을 통하여 비정상 시계열을 정상 시계열로 변환해주었지만, 여전히 남아 있는 오차가 IID나 WN은 아닐 때를 추정하기 위해 시계열 모형이 필요합니다. 우선 어떤 상황에서 어떤 모형을 사용해야 하는지 판단하는 기준인 ACF와 PACF에 대해 배워보겠습니다.

2. 자기상관함수 ACF(Autocorrelation Function)

1주차에서 배운 ACF 함수입니다. 자기상관함수는 시차가 h인 **시계열 간의 상관관계**를 의미하며, 정상성을 만족한다면 시차에만 의존하는 특성이 있습니다. 대표적인 특징은 다음과 같습니다.

🥦 **ACF의 특징**

i. $\rho(0) = 1 \ \ (\because \gamma(0) = var(X_t))$

ii. $\rho(-h) = \rho(h)$

iii. $|\gamma(h)| \le \gamma(0) \ \ for\ all\ h \in Z \to |\rho(h)| \le 1$

3. 부분자기상관함수 PACF(Partial Autocorrelation Function)

PACF는 이름에서도 알 수 있는 것처럼 ACF에 'partial'이 추가된 함수입니다. 'partial'의 의미를 알아보기 위해 **부분상관계수(partial correlation coefficient)**에 대해 알아보겠습니다.

> X를 아이스크림 판매량, Y를 범죄발생건수, Z를 인구 수라고 설정하겠습니다. 만약 인구 수 Z가 시간에 따라 증가한다면 X와 Y 역시 증가하게 되고, 그 결과 X와 Y의 상관관계가 높아질 것입니다. 하지만 이는 두 변수 모두 Z의 영향을 받아 발생한 결과이기 때문에, Z의 영향을 제거한 순수한 X와 Y의 상관계수를 구해야 합니다. 이때 사용하는 개념이 **부분상관계수**입니다.

Z의 효과를 배제한 X와 Y의 부분상관계수는 다음과 같습니다.

$$\rho_{XY,Z} = \frac{E([X - E(X|Z)] * [Y - E(Y|Z)])}{\sqrt{E[X - E(X|Z)]^2 * E[Y - E(Y|Z)]^2}} = \rho(X^*, Y^*)$$

① 조건부 기댓값 $E(X|Z)$: X가 Z에 의해 설명되는 부분

② 조건부 기댓값 $E(Y|Z)$: Y가 Z에 의해 설명되는 부분

③ $X^* = X - E(X|Z)$: X에서 Z의 영향력을 제거하고 남은 부분

④ $Y^* = Y - E(Y|Z)$: Y에서 Z의 영향력을 제거하고 남은 부분

여기까지가 'auto'가 붙지 않은 부분상관계수였습니다. 그렇다면 부분자기상관함수는 자기 자신과의 부분상관계수라고 할 수 있겠죠?

부분자기상관계수(Partial Autocorrelation Coefficient)란 X_t와 X_{t+k}의 상관관계를 구할 때, 그 둘 사이에 존재하는 {$X_{t+1}, X_{t+2}, \ldots, X_{t+k-1}$}의 영향을 제외하고 구한 상관계수입니다. 부분자기상관함수(Partial Autocovariance Function; PACF)는 보통 $\alpha(k)$로 표현하며,다음과 같이 정의합니다.

🥦 **PACF**

i. $\alpha(0) = Corr(X_1, X_1) = 1$

ii. $\alpha(1) = Corr(X_2, X_1) = \rho(1)$

iii. $\alpha(k) = Corr(X_{k+1} - P_k^* X_{k+1}, X_1 - P_k^* X_1),\ k \ge 2$

이때 $P_k^* X_{k+1}$와 $P_k^* X_1$는 각각 X_{k+1}와 X_1의 오차항을 최소화하는 추정치($\{1, X_2, \ldots, X_k\}$가 X_{k+1}와 X_1에 미치는 영향이 가장 많은 추정치)라고 보시면 됩니다.

II. Linear Process (선형 과정)

선형과정은 한 마디로 하면 $\{Z_t\}$~$WN(0, \sigma^2)$들의 선형결합입니다. 이를 식으로 표현하면 다음과 같습니다.

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$$

이때, 선형결합의 계수는 $\sum_j |\psi_j| < \infty$(absolutely summable)의 조건을 만족해야 합니다. 왜 시계열을 선형과정으로 표현하는 것일까요? 다음 네 가지 이유를 들 수 있습니다.

i. 선형과정의 공분산 계산은 매우 간단합니다.
ii. 해석과 추정이 잘 발달되어 있습니다.
iii. 정상 확률 과정의 선형 결합은 또 다시 정상 확률 과정 조건을 만족합니다.
iv. Wold Decomposition에 따르면 모든 약한 의미의 정상확률과정은 선형과정과 결정적 과정으로 나타낼 수 있습니다.

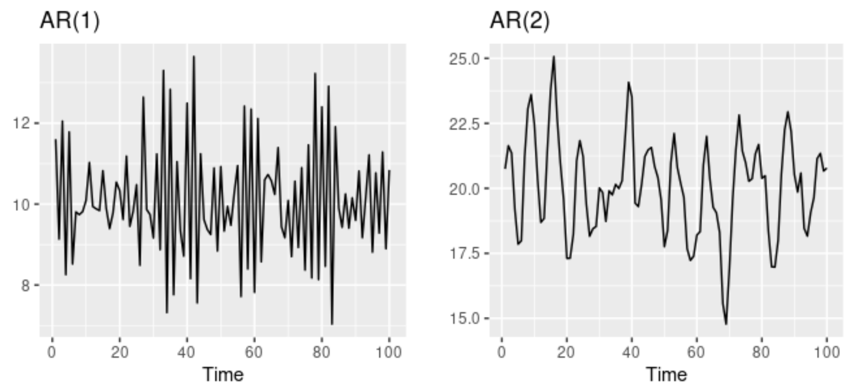
오늘 배울 AR과 MA 모형 또한 선형과정의 모형 중 하나입니다.

III. AR(Auto Regressive Model): 자기회귀 모형

1. 정의

AR 모형이란 현 시점의 관측값을 과거 시점의 관측값과 현 시점의 오차의 함수 형태로 나타내는 모형입니다. 이렇게 관측값이 자기 자신의 과거에 회귀시킨다는 의미에서 자기회귀라는 표현을 사용합니다. 이때 Z_t~$WN(0, \sigma^2)$이며, 아래와 같이 표현할 수 있습니다.

> 🥦 **AR 모형**
>
> $AR(1)$: $\quad X_t = \phi_1 X_{t-1} + Z_t$
>
> $AR(p)$: $\quad X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t$



위 식들은 우리가 흔하게 알고 있는 회귀식과 유사합니다. 이처럼 AR 모형은 관측값을 자기 자신의 과거로 회귀시킨다는 의미에서 '**자기회귀 모형**'이라고 부릅니다.

2. 특성방정식

AR(p)에서 p는 몇 시점 전까지의 관측값의 선형결합으로 표현했는지를 나타냅니다. 1주차에 배웠던 후향연산자(Backshift Operator)를 기억하시나요? 후향연산자를 활용하여 AR(p) 모형을 다음과 같이 표현할 수 있습니다.

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_q X_{t-p} + Z_t \\ &= \phi_1 B X_t + \phi_2 B^2 X_t + \cdots + \phi_p B^p X_t + Z_t \end{aligned}$$

후향 연산자를 이용하여 나타낸 식을 Z_t에 대해 정리하면 다음과 같습니다.

$$Z_t = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t$$

여기서, $(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$를 **특성 방정식(character equation)** $\phi(B)$라고 합니다. 이 특성 방정식을 사용하여 AR 모델을 다음과 같이 표현할 수 있습니다.

$$Z_t = \phi(B) X_t$$

3. AR 모형의 조건

AR 모형은 다음과 같은 두 가지 특성을 만족해야 합니다.

i. **정상성(Stationarity)**: 시계열의 확률적 특성이 시점에 의존하지 않아야 하는 특성

ii. **인과성(Causality)** : t시점의 관측값이 과거시점의 오차항으로 설명될 수 있다는 특성

$$\psi_j = 0,\ j < 0 \ <-> X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$$

결론부터 말씀드리면, AR 모형은 $|\phi| < 1$일 때 인과성과 정상성을 만족합니다. 또한 이는 '$\phi(B) = 0$의 근의 절댓값이 1보다 커야 한다'와 동치입니다. 왜 $|\phi| < 1$라는 조건이 AR 모형이 이 두 가지 특성을 만족하는데 필요한지 가장 간단한 AR(1) 모형($X_t = \phi_1 X_{t-1} + Z_t$)을 통해 알아보겠습니다.

$$\begin{gathered} X_t = \phi_1 X_{t-1} + Z_t \\ = \phi_1(\phi_1 X_{t-2} + Z_{t-1}) + Z_t = \phi_1^2 X_{t-2} + \phi_1 Z_{t-1} + Z_t \\ = \phi_1^2(\phi_1 X_{t-3} + Z_{t-2}) + \phi_1 Z_{t-1} + Z_t = \phi_1^3 X_{t-3} + \phi_1^2 Z_{t-2} + \phi_1 Z_{t-1} + Z_t \\ \vdots \\ = \phi_1^{M+1} X_{t-M-1} + \sum_{j=0}^{M} \phi_1^j Z_{t-j} \end{gathered}$$

AR(1) 식은 위와 같이 과거 시점의 관측값과 오차들의 선형결합으로 정리할 수 있습니다. 인과성을 만족하기 위해서는 오차항만으로 관측값을 설명해야 합니다.

[$|\phi < 1|$인 경우]

$$X_t = \phi_1^{M+1} X_{t-M-1} + \sum_{j=0}^{M} \phi_1^j Z_{t-j}$$

위 식에서 $M \to \infty$ 이면 $\phi_1^{M+1} X_{t-M-1} \to 0$으로 수렴, 나머지 부분은 $\sum_{j=0}^{M} \phi_1^j Z_{t-j}$이 됩니다. 즉 $|\phi| < 1$인 경우에는 오차항의 선형결합(Z_t)만 남아 인과성을 만족합니다. 또한 정상 시계열의 선형결합은 여전히 정상 시계열임을 확인하였기 때문에 정상성 역시 만족함을 알 수 있습니다.

*$|\phi| = 1$인 경우와 $|\phi| > 1$인 경우는 Appendix를 참고해주세요 ~

4. ACF

AR(1) 모형을 이용하여 ACF를 계산해보겠습니다. 계산의 편의를 위해 $E(X_t) = 0$을 가정하겠습니다.

[1] ACF 식을 유도하기 위해 AR(1) 식 양변에 X_{t-h}를 곱해줍니다.

$$\begin{gathered} X_t = \phi_1 X_{t-1} + Z_t \\ X_t X_{t-h} = \phi_1 X_{t-1} X_{t-h} + Z_t X_{t-h} \end{gathered}$$

[2] 위 식에 기댓값을 취해 ACF 식을 구합니다.

$$\begin{gathered} \gamma(h) = \phi_1 \gamma(h-1) + Cov(Z_t, X_{t-h}) = \phi_1 \gamma(h-1) \\ (\because Cov(Z_t, X_{t-h}) = Cov(Z_t, \phi_1^{M+1} X_{t-M-1} + \sum_{j=0}^{M} \phi_1^j Z_{t-j}) = 0) \\ \gamma(h) = \phi_1(\phi_1 \gamma(h-2)) = \cdots = \phi_1^h \gamma(0) \\ \frac{\gamma(h)}{\gamma(0)} = \phi_1^h = \rho(h) \end{gathered}$$

3.에서 정상성을 만족하기 위해 $|\phi| < 1$가 되어야 함을 확인했기 때문에, h가 커짐에 따라 AR 모형의 ACF는 지수적으로 감소함을 알 수 있습니다.

5. PACF

AR(p) 모형식을 이용해 AR 모형의 PACF에 대해 알아보겠습니다. AR(p)는 X_{k+1}을 p 시점 이전 값들로만 표현하는 것이며, 식은 아래와 같습니다.

$$\hat{X_{k+1}} = \phi_1 X_k + \phi_2 X_{k-1} + \cdots + \phi_p X_{k+1-p}$$

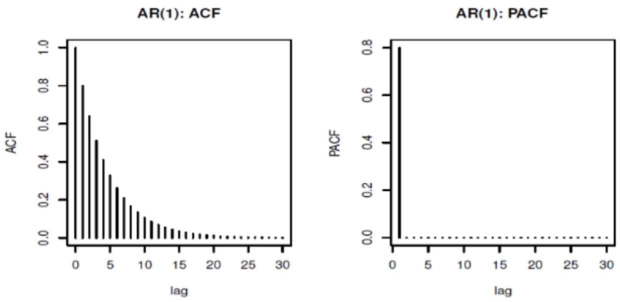
PACF를 유도하기 위해 위의 식을 조금 더 자세히 풀어서 써보면, 아래와 같이 나타낼 수 있습니다.

$$\hat{X_{k+1}} = \phi_1 X_k + \phi_2 X_{k-1} + \cdots + \phi_p X_{k+1-p} + 0X_{k-p} + \cdots + 0X_1$$

이 식과 PACF의 성질을 이용해 AR 모형의 PACF를 정리하면 아래와 같습니다.

$$\alpha(0) := 1,\ \alpha(p) = \phi_p,\ \alpha(k) = 0\ \ if\ \ k > p$$

즉, AR 모형의 PACF는 p 시차 전까지만 존재하며, p 이후로는 모두 0이 됩니다. 이를 'AR(p) 모형의 PACF는 시차 p 이후에 절단된다'라고 표현합니다.



그림을 통해 AR 모형의 ACF와 PACF를 살펴보면, 위에서 확인한 바와 같이 **ACF는 지수적으로 감소, PACF는 p 이후(위 그림에서 p=1) 에 절단**된 양상을 보이고 있음을 알 수 있습니다.

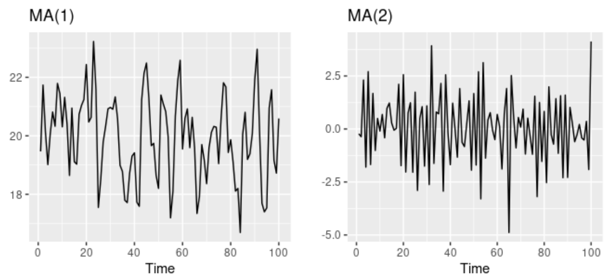
IV. MA(Moving Average Model): 이동평균모형

1. 정의

MA 모형은 관측값을 과거시점의 오차항만을 이용해 설명하는 모델입니다. AR 모형과 마찬가지로 Z_t는 Z_t~$WN(0, \sigma^2)$를 만족하는 백색잡음입니다. 식은 다음과 같습니다.

🥦 **PACF**

$MA(1) : X_t = Z_t + \theta_1 Z_{t-1}$
$MA(q) : X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q X_{t-q}$



2. 특성방정식

$$\begin{aligned} X_t &= Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} \\ &= Z_t + \theta_1 B Z_t + \theta_2 B^2 Z_t + \cdots - \theta_q B^q Z_t \\ &= (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t \end{aligned}$$

위 식에서 $(1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$을 MA(q)의 특성 방정식이라고 합니다. 따라서 MA(p)를 특성 방정식을 이용해 표현하면 다음과 같습니다.

$$X_t = \theta(B) Z_t$$

지금까지 배운 MA(q)를 표현하는 식들을 정리하면 다음과 같습니다.

3. MA 모형의 조건

MA 모형은 백색잡음이라고 가정하는 오차들의 선형결합, 그 중에서도 t시점 이전의 오차항들로 표현하는 모델이기 때문에 **정상성과 인과성을 자체적으로 만족**합니다. 이에 더하여, **가역성** 조건 역시 만족해야 합니다.

i. **가역성(Invertibility)**

: t시점의 오차항을 과거 시점의 관측값으로 표현할 수 있는 성질입니다. 이러한 성질은 예측에 유용합니다. $\sum_{j=0}^{\infty} |\pi_j| < \infty$인 {$\pi_j$ }가 존재할 때, 다음의 식을 만족하면 확률 과정이 invertible하다고 할 수 있습니다.

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \ for\ all\ t$$

즉, Z_t에 대한 식을 X_t로 표현할 수 있는지 여부를 따지는 조건이라고 이해하시면 됩니다. 그렇다면, 가역성은 어떤 조건에서 만족하는지 MA(1) 모형을 통해 알아보겠습니다.

$$\begin{gathered} X_t = Z_t + \theta Z_{t-1} = (1 + \theta B) Z_t \\ (1 + \theta B)^{-1} X_t = Z_t \\ (1 + \theta B)^{-1} = \frac{1}{1 - (-\theta B)} = 1 - \theta B + (\theta B)^2 - (\theta B)^3 + \ldots \end{gathered}$$

위와 같이 MA(1) 식을 X_t에 대해 정리한 후 식을 풀어보면 무한등비급수의 합 형태로 표현할 수 있습니다. 따라서 가역성 조건 역시 $|\theta| < 1$일 때만 성립하며, 이는 특성방정식 $\theta(B) = 0$의 근의 절댓값이 1보다 커야 한다는 조건과 동치입니다.

4. ACF

AR 모형에서와 같이, 양변에 X_{t-h}를 곱하고, 기댓값을 취하여 ACF를 구해보겠습니다.

[MA(1) 모형]

$$X_{t-h}X_t = X_{t-h}(Z_t + \theta_1 Z_{t-1}) = X_{t-h}Z_t + \theta_1 X_{t-h}Z_{t-1}$$
$$\gamma(h) = Cov(X_{t-h}, Z_t + \theta_1 Z_{t-1}) = Cov(Z_{t-h} + \theta_1 Z_{t-h-1}, Z_t + \theta_1 Z_{t-1})$$

i. $h = 0$인 경우

$$\gamma(0) = Cov(Z_t + \theta_1 Z_{t-1}, Z_t + \theta_1 Z_{t-1}) = \sigma^2 + \theta_1^2\sigma^2 = (1+\theta_1^2)\sigma^2$$

ii. $h = 1$인 경우

$$\gamma(1) = Cov(Z_{t-1} + \theta_1 Z_{t-2}, Z_t + \theta_1 Z_{t-1}) = \theta_1\sigma^2$$

iii. $h > 2$인 경우

$$\gamma(h) = Cov(Z_{t-h} - \theta_1 Z_{t-h-1}, Z_t - \theta_1 Z_{t-1}) = 0$$

정리하면 다음과 같습니다. (MA(q))

$$\rho(k) = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\theta}{1+\theta^2},\ k = 1 \\ 0,\ k \geq 2 \end{cases}$$

따라서 MA(q) 모형의 ACF는 시차 q 이후 절단된다는 것을 알 수 있습니다.

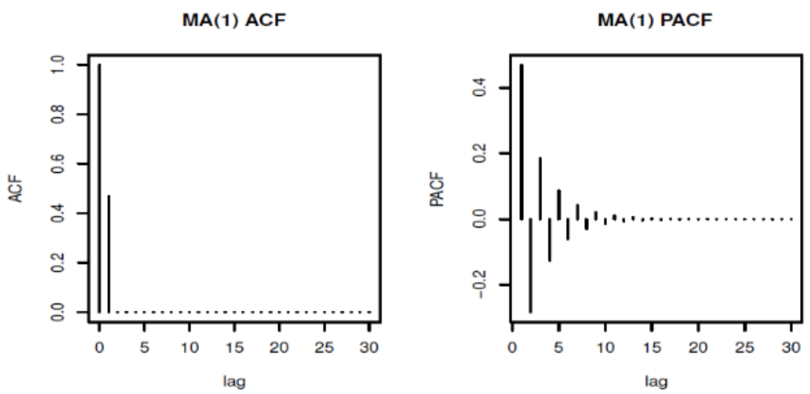
5. PACF

MA 모형에서의 PACF는 어떤 패턴을 보이는지 알아보겠습니다.

$$\phi_{kk} = \frac{-\theta^k(1-\theta^2)}{1-\theta^2(k+1)}, k \geq 1$$

위 식은 MA(1) 모형의 PACF를 계산한 결과입니다. 5.3에서 MA 모형은 $|\theta| < 1$ 일 때 성립함을 배웠으므로, 위 식은 **시차 k가 커질수록 0에 수렴**하게 됩니다. 즉, MA 모형에서는 PACF가 지수적으로 감소함을 알 수 있습니다.

위의 MA 모델의 PACF 는 Crammer 공식을 사용하여 얻어진 결과입니다. 다만 계산 과정을 자세히 다루지는 않겠습니다~



MA(1) 모형의 ACF와 PACF 그래프입니다. 위에서 공식을 통해 확인한 것과 같이 **ACF는 시차 1까지 존재하며, PACF는 점점 0으로 수렴**하고 있음을 알 수 있습니다.

VI. ARMA(p,q)

ARMA 모형은 이름 그대로 AR 모형과 MA 모형을 동시에 포함하는 모형입니다. 그렇다면 왜 두 모형을 통합한 모형이 또 필요한 것일까요? 이는 **“모수의 절약(parsimony)”**과 관련이 있습니다. 시계열 자료를 순수한 자기회귀모형이나 이동평균모형으로만 설명하려면 p나 q의 값이 너무 커질 가능성이 있습니다. 이때, 추정해야 할 모수의 개수가 많아지면 일반적으로 추정의 효율성이 떨어지고 해석이 어렵다는 문제가 발생합니다. 이러한 문제를 해결하기 위해 ARMA 모형을 사용하는 것입니다! 그럼 지금부터 ARMA 모형에 대해 자세히 알아보겠습니다.

1. 정의

앞서 설명한 바와 같이 ARMA 모형은 AR 모형과 MA 모형을 동시에 포함하는 모형이며, ARMA(p,q)의 식은 다음과 같습니다.

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$$

2. 특성방정식

$$\begin{gathered} X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q} \\ X_t - \phi_1 B X_t - \phi_2 B^2 X_t - \cdots - \phi_p B^p X_t = Z_t + \theta_1 B Z_t + \theta_2 B^2 Z_{t-2} + \cdots + \theta_q B^q Z_t \\ (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t \\ \phi(B) X_t = \theta(B) Z_t \end{gathered}$$

왼쪽 식은 AR 모형의 특성방정식이며, 오른쪽 식은 MA 모형의 특성방정식입니다!

3. ARMA 모형의 조건

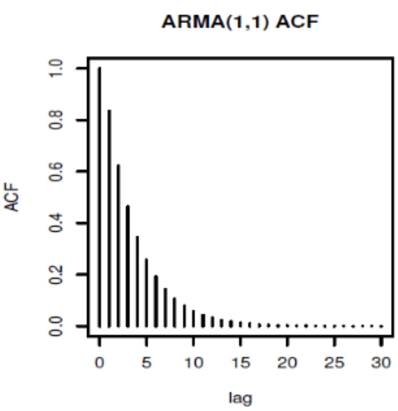
ARMA 모형은 정의에서 유추할 수 있는 것처럼 AR 모형의 조건과 MA 모형의 조건을 모두 만족해야 합니다. 즉, 정상성과 인과성, 가역성을 모두 만족해야 합니다. 추가적으로, ARMA 모형은 **식별성(Identifiability)** 조건까지 만족해야 합니다. 여기서 식별성이란, 주어진 파라미터 조합에 대해 단 하나의 모형이 대응되는 특성을 의미합니다. 식을 통해 확인해보겠습니다.

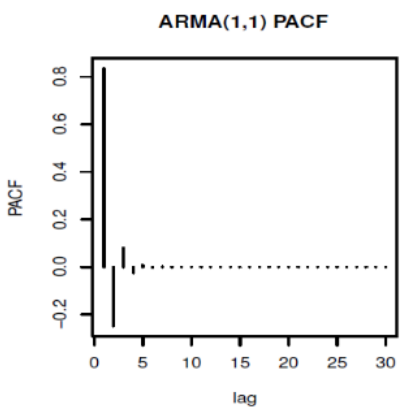
$$X_t = Z_t$$

위 식은 $(1-\phi B) = (1+\theta B) <=> \phi = -\theta$를 만족하는 ARMA(1,1) 모형입니다. 하지만 동시에 WN이기도 하기에, 정확히 어떤 모형인지 파악하지 어렵습니다. 이처럼 정상성, 인과성, 가역성만으로는 정확한 모형을 식별할 수 없는 문제가 발생하기 때문에 $\phi + \theta \neq 0$의 식별성 조건이 필요한 것입니다.

4. ACF와 PACF

ARMA 모형의 ACF 그래프와 PACF 그래프는 **모두 지수적으로 감소하거나 sin함수 형태로 소멸되는 양상**을 보입니다. AR(p) 모형은 PACF가 p+1차부터 절단되었고, MA(q) 모형은 ACF 그래프가 q+1차부터 절단되는 특징이 있었습니다. 하지만 ARMA 모형은 두 그래프 모두 절단되지 않고 지수적으로 감소하기 때문에, 모형 식별을 위한 추가적인 방법이 필요합니다. 해당 방법은 모형 식별 부분에서 다루도록 하겠습니다.





VII. 정상 시계열 모형의 적합 절차

지금까지 WN이 아닌 정상 시계열의 오차를 설명하기 위한 모형인 AR, MA, ARMA에 대해 배웠습니다. 그렇다면 언제, 어떤 모형을 사용해 시계열 자료를 설명해야 할까요? 지금부터 상황에 맞는 모델을 선택하기 위한 모형 적합 절차에 대해 알아보겠습니다!

1. 모형 식별

가장 먼저 사용할 모형과 해당 모형의 차수를 결정해야 합니다. AR(p) 또는 MA(q)의 경우, 시계열 자료의 ACF와 PACF 그래프를 확인해보았을 때, 그래프가 절단되는 특징이 나타나기 때문에 p 또는 q의 차수를 쉽게 결정할 수 있습니다. 하지만 앞서 살펴봤듯이, ARMA 모형의 경우 ACF와 PACF 모두 지수적으로 감소하기 때문에 이 그래프만을 활용하여 모형을 식별하는 데에는 한계가 있습니다. 따라서 AIC, AICC, BIC 등의 Information Criteria를 사용하게 됩니다.

$$Information\ Criteria = \{goodness\ of\ fit + model\ complexity\}$$

이러한 IC는 모형의 복잡도를 함께 고려하기 때문에, 모형 식별을 위한 적절한 지표라고 볼 수 있습니다. 여러 모형들 중 가장 작은 IC 값을 가지는 모형을 선택하면 됩니다.

> 🥦 **Information Criteria**
> i.
> **AIC**(Akaike Information Criteria): $2\ln L_n(\hat{\theta}) + 2(p+q+1)$
> ii.
> **AICC** (AIC biased corrected): $2\ln L_n(\hat{\theta}) + \frac{2(p+q+1)n}{n-(p+q+1)+1}$
> iii.
> **BIC** (Bayesian Information Criterion): $2lnL_n(\hat{\theta}) + (p+q+1)\ln n$

위 식에서 $-2\ln L_n(\hat{\theta})$는 SSE(Sum of Squared Error)에 대한 approximation을 나타냅니다.

2. 모수 추정

모수의 차수를 결정했다면, 그 다음으로는 모수를 추정해야 합니다.

> 🥦 **모수추정법**
> i.
> **최대가능도추정법(MLE)**: 관측된 시계열의 결합확률밀도함수인 모수의 가능도 함수 (likelihood function)를 최대화하는 모수의 추정량을 구하는 방법
> ii.
> **최소제곱법(LSE)**: 오차의 제곱합이 가장 작게 되도록 하는 모수의 추정량을 구하는 방법
>
> iii. 적**률추정법 (MME/MoM)** : 모집단의 적률을 상응하는 표본의 적률로 대체한 후, 방정식을 풀어 모수의 추정량을 구하는 방법

보편적으로 MLE의 성능이 가장 좋다고 알려져 있지만, MLE를 구하기 위해서는 수학적 최적화가 필요합니다. 효과적인 접근을 위해 LSE와 MME를 MLE에 대한 초기값으로 사용하여 MLE를 추정하기도 합니다.

3. 모형 진단

모형을 식별하고 모수 추정까지 마쳤다면, 모형이 적합한지 진단해야 합니다. 모형 진단은 모수에 대한 검정과 잔차에 대한 검정으로 진행됩니다.

i. **모수에 대한 검정**

- 모형의 조건을 만족하는지 확인: 정상성과 가역성 조건 만족 여부 (모수의 절댓값이 1보다 작은지), 식별성 만족 여부 (모수의 합이 0이 아닌지를 확인)
- 모수의 유효성 확인: 모수 ≠ 0인지 확인

ii. **잔차에 대한 검정**

- 추세, 계절성, 이상치가 없는지 확인
- WN(0, σ^2)을 따르는지 확인 (자기상관성 유무): 잔차에 대한 ACF, PACF 그래프 / Ljung-Box test/ McLeod-Li test / Different sign test
- 정규성을 만족하는지 확인 : 잔차의 QQ plot / Jarque-Bera test

4. 예측

모형 식별, 파라미터 추정, 모형 진단까지 마쳤다면, 모형을 선택하여 예측을 진행할 수 있습니다.
예측에는 과거의 모든 정보를 알고 있다고 가정하는 infinite한 방법과, 알고 있는 자료를 사용해 예측하는 finite한 방법이 있습니다. 클린업에서는 실제로 주로 사용되는 finite한 방법을 알아보도록 하겠습니다!

$$P_n X_{n+h} = a_0 * 1 + a_1 X_n + a_2 X_{n-1} + \cdots + a_n X_1$$

위 식은 개의 자료를 이용해 시점을 예측했음을 의미합니다. 즉, 가지고 있는 데이터의 **선형결합**을 활용해 미래를 예측하는 방법입니다.

계수 $\{a_0, a_1, ..., a_n\}$는 MSPE(Mean Squared Prediction Error)를 최소화하는 방향으로 추정합니다.

$$MSPE = E[X_{n+h} - P_n X_{n+h}]^2$$
$$= E[X_{n+h} - (a_0 * 1 + a_1 X_n + a_2 X_{n-1} + \cdots + a_n X_1)]^2$$

계산 방식은 회귀분석에서의 LSE와 동일합니다! 위의 MSPE 식을 각각의 계수 $a_0, a_1, ..., a_n$에 대해 미분한 다음, 미분한 식=0으로 놓고 각 normal equation을 연립하여 해를 구하면 됩니다. (하지만 n개의 연립방정식을 직접 계산하여 해를 찾는 것은 사실상 불가능하기 때문에, 컴퓨팅의 힘을 빌리게 됩니다 ㅎㅎ.)

> 지금까지 정상시계열을 다루는 모형들에 대해 배워보았습니다. 즉, 모두 정상화를 거친 데이터에 적용하는 모형들이었습니다. 지금부터는 정상화를 진행하지 않은 원본 데이터에 적용할 수 있는 **비정상 시계열 모형**을 배워보겠습니다!

VIII. ARIMA(Autoregressive Integrated Moving Average Model)

1. 정의

이번에 배울 모형은 비정상 시계열을 다루는 모형입니다. ARIMA 모델은 이름에서 확인할 수 있듯이 차분과 ARMA가 결합된 형태입니다. d차 차분 후, $Y_t = (1-B)^d Z_t$가 정상과정 ARMA(p,q)를 따를 때 $\{X_t\}$는 **자기회귀누적이동평균과정 ARIMA(p,d,q)**를 따른다고 합니다. 다음과 같이 표현할 수 있습니다.

$$\phi(B)(1-B)^d X_t = \theta(B) Z_t$$

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)(1-B)^d X_t = (1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q) Z_t$$

$\phi(B)$는 AR의 특성방정식, $\theta(B)$는 MA의 특성방정식, $(1-B)^d$는 d차 차분을 의미합니다.

식에서도 알 수 있듯이, d=0인 ARIMA(p,0,q) 모형은 ARMA(p,q) 모형과 동일하며, d ≥ 1인 경우 d차의 추세를 제거하게 됩니다. 따라서 ARIMA 모델은 주어진 시계열 데이터가 **polynomial trend를 가지고 있고, 그 오차가 ARMA(p,q)를 따를 때** 사용하는 모델입니다. 앞서 배운 모델들과는 달리, 모형 자체에 차분을 포함하여 상대적으로 간결하게 처리할 수 있으며, 여전히 선형 과정을 따르기 때문에 예측에 용이하다는 장점을 지닙니다.

🥦 **(참고) 차분(differencing)과 ARMA를 결합한 모형을 ARDMA가 아닌 ARIMA라고 정의한 이유**

ARIMA에서 I는 누적(integration)을 의미합니다. 아래 식을 통해 확인해보겠습니다!

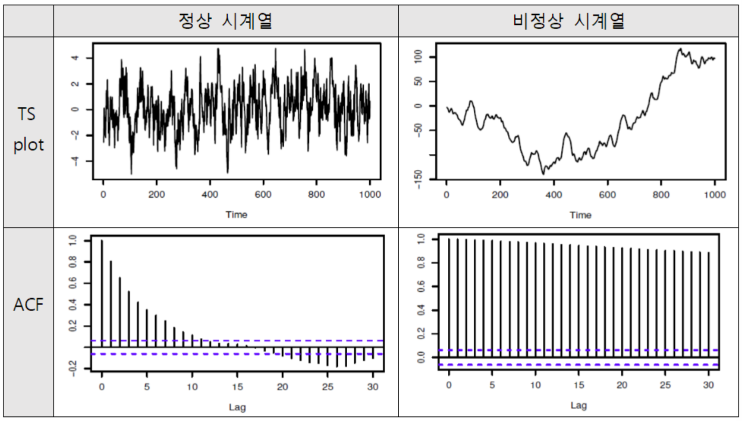
$$\phi(B)(1-B)X_t = \theta(B)Z_t$$
$$Y_t = (1-B)X_t = X_t - X_{t-1}$$

$$X_t = X_{t-1} + Y_t = (X_{t-2} + Y_{t-1}) + Y_t = \cdots = X_0 + \sum_{j=1}^{t} Y_j$$

위 식과 같이 X_t는 Y_t의 누적합으로 볼 수 있으며, 이는 random walk의 확장으로 해석할 수 있습니다. 따라서 ARIMA 모형을 자기회귀누적이동평균모형이라고 부릅니다.

2. 모델 적합 절차

1) 시계열 플랏과 ACF 그래프를 통해 정상 시계열인지 비정상 시계열인지 판단합니다.



위 그림처럼 정상 시계열은 ACF가 지수적으로 감소하지만, 비정상 시계열은 ACF가 천천히 감소하는 것을 이용해 적절한 모형을 선택할 수 있습니다.

2) 비정상 시계열에서 추세가 관측된다면, 차분을 통해 정상화를 진행합니다.

관측되는 추세에 맞추어 d차 차분을 적용합니다. 이때 차분의 d가 1,2를 넘어가면 과대차분의 위험이 있기에 주의합니다.

과대차분이란 말 그대로 차분을 과하게 하는 것입니다. 이미 정상화가 되어있음에도 불구하고 추가적으로 차분하게 된다면, 정상성 자체는 여전히 만족하지만 ACF가 복잡해지거나 분산이 커지며, 불필요한 상관관계를 생성해 적합 과정이 보다 복잡해질 수 있는 문제가 발생합니다. 보통 과대차분을 방지하기 위해 1차 또는 2차 차분을 많이 사용합니다.

비정상 시계열 데이터에 1차 또는 2차 차분을 적용해보고, ACF와 PACF 그래프를 확인했을 때 지수적으로 감소하는 그래프가 나타난다면 적절하다고 볼 수 있습니다.

3) 모형 적합 절차에 따라 p, q의 차수를 결정하고, 모수를 추정하고 진단까지 마친 모형을 통해 예측을 진행합니다.

지금까지 다항 **추세가 존재하는 비정상 데이터**에 ARMA를 적용하는 ARIMA모형에 대해 배웠습니다. 그렇다면 **추세와 계절성이 모두 존재하는 경우**에는 어떤 모형을 사용할까요?

IX. SARIMA(Seasonal Autoregressive Integrated Moving Average Model)

SARIMA는 시계열 자료에 추세와 계절성이 모두 있을 때 사용할 수 있는 모형입니다.

1주차에 배운 정상화에서의 계절성은 결정적(deterministic) 계절성으로, **모든 주기에서 계절성이 동일함**을 가정했습니다. 하지만 현실에서는 모든 계절 성분이 결정적이지는 않으며, 다른 성분들과의 상관관계가 존재할 수 있습니다. 이렇게 주기가 변함에 따라 계절성도 변화

할 수 있음을 반영하여 **확률적 분석**을 하는 모형이 SARIMA 모형입니다.

1. 정의

SARIMA 모형은 Seasonal ARIMA의 줄임말로, **추세와 계절성이 모두 존재하는 비정상 데이터**에 적용할 수 있는 모형입니다. 가장 확장된 형태의 ARMA모형으로, 계절성 사이의 상관관계에 대해 모델링하는 확률적 접근법으로 이해하시면 됩니다!

	month 1	month 2	···	month 12
Year 1	Y_1	Y_2	...	Y_{12}
Year 2	Y_{13}	Y_{14}	...	Y_{24}
⋮	⋮	⋮	⋮	⋮
Year r	$Y_{1+12(r-1)}$	$Y_{2+12(r-1)}$	...	$Y_{12+12(r-1)}$

다음과 같이 주기 s=12인 시계열을 정리해보면, column을 고정했을 때 이 시계열들이 ARMA(P,Q)를 따른다고 생각해보겠습니다. 그때, 다음과 같이 표현할 수 있습니다. 여기서 U_t~$WN(0, \sigma_U^2)$입니다.

$$\begin{gathered} Y_{j+12t} - \Phi_1 Y_{j+12(t-1)} - \cdots - \Phi_p Y_{j+12(t-P)} \\ = U_{j+12t} + \Theta_1 U_{j+12(t-1)} + \cdots + \Theta_Q U_{j+12(t-Q)}, \\ t = 0, 1, \ldots, 11 \\ <=> \Phi(B^{12}) Y_t = \Theta(B^{12}) U_t \end{gathered}$$

위 식이 어떤 의미인지 자세히 살펴보면, month가 달라져도 파라미터는 같습니다. 위의 오차항 U_t를 백색잡음이 아닌 ARMA(p,q)를 따르는 시계열이라고 생각해봅시다. Month가 달라져도 이들 간 correlation이 있을 수 있기 때문에 U_t 역시 다음과 같이 표현할 수 있습니다.

$$\phi(B)U_t = \theta(B)Z_t, Z_t \sim WN(0, \sigma^2)$$

앞선 두 과정을 합쳐서 표현하면 다음과 같습니다.

$$\begin{gathered} \phi(B^{12})Y_t = \Theta(B^{12})\phi^{-1}(B)Z_t \\ \phi(B)\Phi(B^{12})Y_t = \theta(B)\Theta(B^{12})Z_t, Z_t \ \sim WN(0, \sigma^2) \end{gathered}$$

뿐만 아니라, 추세가 존재할 경우 d차 차분까지 포함할 수 있습니다.

$$Y_t = (1-B)^d (1-B^{12}) X_t$$

따라서 SARIMA(p,d,q)x(P,D,Q) 모델은 다음과 같이 표현할 수 있습니다.

$$\phi(B)\Phi(B)^{12}(1-B)^d(1-B)^{12})^D Xt = \theta(B)\Theta(B^{12})Zt, Z_t \sim WN(0, \sigma^2)$$

정리해보자면 SARIMA는 추세와 계절성을 동시에 포함할 수 있는 모형이며 SARIMA모형에는 순수SARIMA와 승법SARIMA 두 가지가 존재합니다. 각각의 경우를 더 자세하게 살펴보겠습니다. 방금 설명한 모형은 승법 SARIMA에 해당합니다.

2. 순수 SARIMA

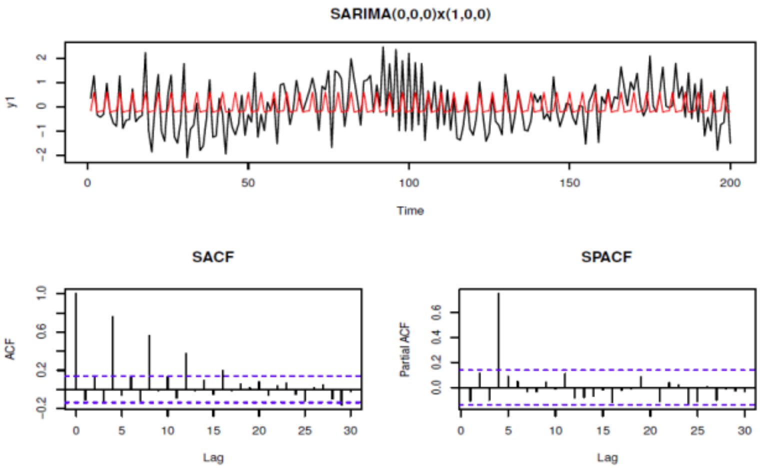
순수SARIMA 모델은 계절성만을 고려하는 모델입니다. 즉, 주기에 대한 차수인 (P,D,Q)s만 있고 전체 시계열에 대한 차수인 (p,d,q)는 없는 모형입니다. 즉, P개의 과거 관측치와 D번의 계절차분, Q개의 과거 오차항으로 현재 관측치를 설명하는 모형입니다. s가 주기라고 할 때, 순수 SARIMA는 아래와 같이 나타낼 수 있습니다.

$SARIMA(0,0,0)(P,D,Q)_s$모형

$$\Phi(B^s)(1-B^s)^D X_t = \Theta(B^s)Z_t$$
$$\Phi(B^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{Ps})$$
$$\Theta(B^s) = (1+\Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs})$$
$$Z_t \sim WN(0,\sigma^2)$$

ACF, PACF

$$(1-.8B^4)X_t = Z_t.$$



위의 그림처럼 순수 SARIMA모형에서의 ACF/PACF는 s에 해당되는 계절 주기에서만 0이 아니고 다른 시차에서는 0이 됩니다. (즉, 모든 시차에 대한 ACF/PACF값을 보는 것이 아닌, 주기에 해당하는 시차의 ACF/PACF를 집중해서 관측해야 한다는 의미입니다. 위 그림은 주기 $s=4$인 경우이며, ACF와 PACF가 시차 4, 8, 12, 16 등에서 유의미하게 나타나는 것을 관측할 수 있습니다.)

그 외의 패턴은 ARMA와 동일합니다. SACF는 감소하는 모양이고 SPACF는 첫번째 주기 이후 절단된 모형입니다. 따라서 위 SACF와 SPACF를 통해 SAR(1,0,0)모형임을 알 수 있습니다.

순수 SARIMA모형은 비계절 요소에 대한 부분은 고려하지 않기에 사용이 제한적입니다. 따라서 비계절 요소를 고려할 때 아래의 승법 SARIMA모형을 사용할 수 있습니다.

3. 승법 SARIMA

승법 SARIMA모형은 계절 이외의 요소의 연관성도 고려하는 모형입니다. ("승법"이라는 단어는 ARIMA×순수 SARIMA의 형태이기 때문입니다.) 순수 SARIMA의 경우 오차는 백색잡음이었지만, 승법 SARIMA의 경우 오차가 ARMA (혹은 ARIMA)를 따릅니다.

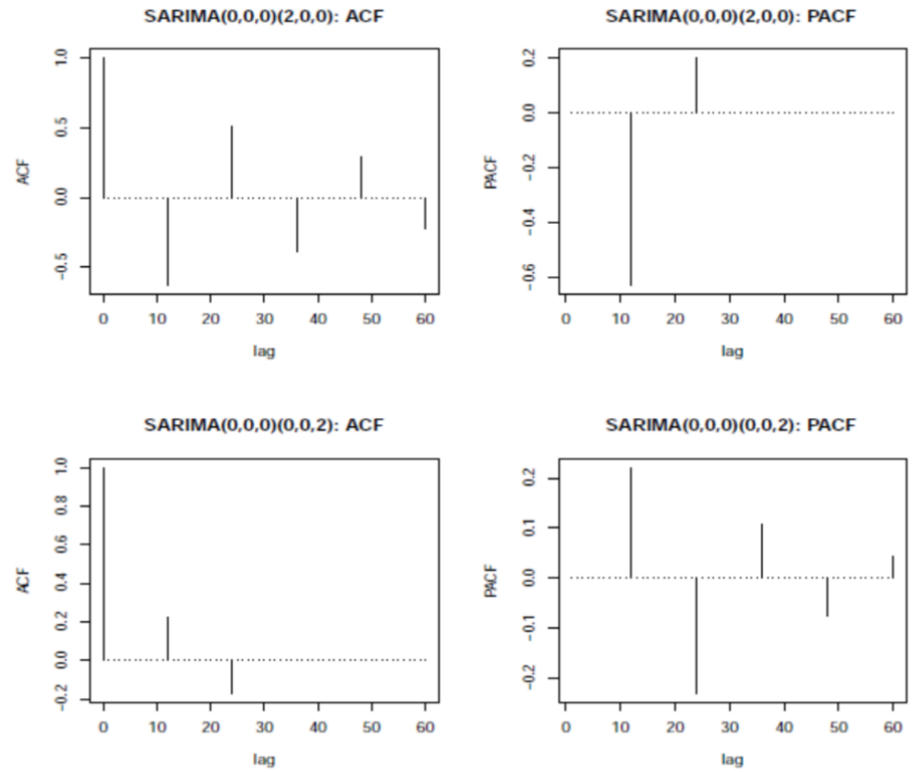
$SARIMA(p,d,q)(P,D,Q)_s$모형

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t$$
$$\Phi(B^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{Ps})$$
$$\Theta(B^s) = (1+\Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs})$$
$$\phi(B) = (1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$$
$$\theta(B) = (1+\theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q)$$
$$Z_t \sim WN(0,\sigma^2)$$

모수절약의 원칙에 따라 순수 SARIMA보다 더 자주 사용됩니다.

ACF, PACF

승법 SARIMA의 경우 비계절성 요소에 관하여 한 주기 내의 ACF와 PACF의 패턴을 통해 파악할 수 있으며 계절적 요소의 경우는 순수 ARIMA와 동일합니다.



지금까지 비정상 데이터에 적용하는 비정상 시계열 모형에 대해 알아봤습니다. 비정상 시계열 모형은 앞에서 언급한 것과 같이 정상화 과정을 따로 진행하지 않아도 된다는 장점이 있습니다. 따라서 예측 결과 역시 원본 데이터에 대한 최종 예측값으로, 추세와 계절성을 따로 더하지 않아도 됩니다!

X. 이분산 시계열 모형

우리가 지금까지 배운 시계열 모형들은 주로 평균 부분의 움직임에 관심을 갖는 모형이었습니다. 전통적인 시계열 자료의 분석에서는 시간에 따른 분산의 변화가 없는 모형을 생각합니다. 오차항이 $\{\epsilon_t\} \sim WN(0, \sigma^2)$인 AR(1) 모형의 조건부 평균과 분산은 다음과 같이 시간에 따라 변화가 없습니다.

$$y_t = \phi_1 y_{t-1} + \epsilon_t$$
$$E(y_t|y_{t-1}) = \phi y_{t-1} \;\; Var(y_t|y_{t-1}) = \sigma^2$$

하지만 수익률, 주가, 환율 등 금융 관련 시계열에서는 분산이 과거 자료에 의존하는(즉, **시간에 따른 이분산성**) 특성을 갖고 있으며, 이러한 이분산성에 관심을 갖고 분석을 진행합니다. **이분산(heteroskedasticity) 시계열 모형**이란 조건부 분산을 시간의 함수로 표현하는 시계열 모형입니다. 시간에 따른 이분산성을 경제학 분야에서는 위험을 측정하는 수단으로서 **변동성(volatility)**이라는 용어를 많이 사용하며, 이는 통계학에서의 **조건부 분산(conditional variance)**입니다. 즉, 이분산 시계열모형은 조건부 분산을 시간의 함수로 표현하는 시계열 모형인 것입니다.

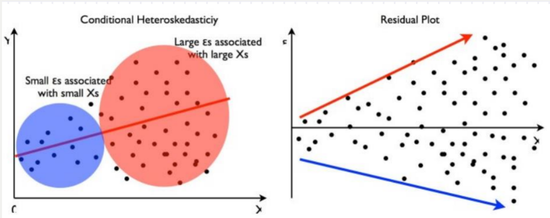
🥦 **이분산성(heteroscedasticity)**

이분산성이란, 분산이 시간이나 관측치별로 독립적으로 나타나지 않거나 일정하지 않음을 의미합니다.

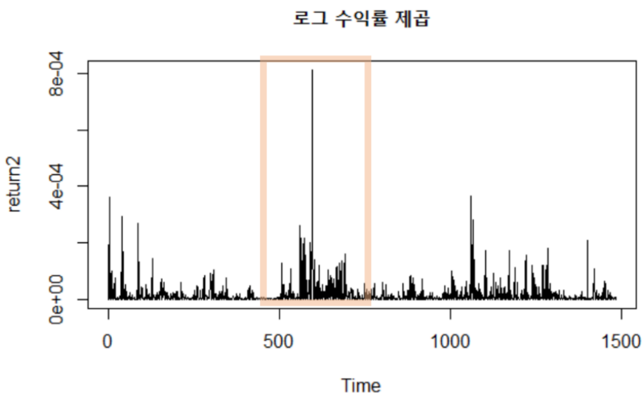
이분산성에는 조건부 이분산성과 비조건부 이분산성이 있습니다.

i. 비조건부 이분산성(uncoditional heteroskedasticity)
→ 일반적인 구조적 변동성의 변화가 이전 기간의 변동성과 관련이 없을 때

ii. 조건부 이분산성(conditional heteroskedasticity)
→ 일정하지 않은 변동성이 이전 시점의 변동성과 연관이 있을 때



1. 변동성 집중(volatility clustering)

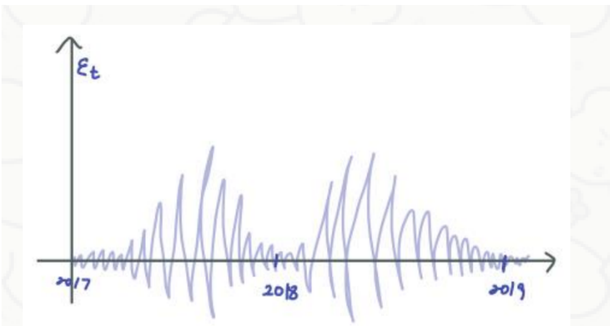


변동성 집중이란 위의 그림과 같이 한 번 나타난 큰 변화가 당분간 계속 큰 변화를 유지하며, 작은 변화는 당분간 지속적으로 작은 변화를 유지하는 경향을 의미합니다. 이러한 변동성 집중에 의한 변화 폭은 $\sqrt{\ }$나 log 등의 자료변환으로 상쇄되지 않으며, 오차의 제곱에 대해 적절한 시계열 모형을 적용하여 설명할 수 있습니다.

2. ARCH(Auto-Regressive Conditional Heteroscedasticity) : 자기회귀이분산모형

ARCH 모델의 흐름은 다음과 같습니다.

우리는 지금 영화관 매표 기록에 대한 데이터를 모델에 적합시키려 한다고 생각해봅시다. 먼저, 최적의 모델을 찾아 적합시킨 뒤 오차항만 남았다고 생각해보겠습니다. 오차항을 이용해 플랏을 그렸더니 다음과 같이 나왔습니다.



오차항이 변할 땐 크게 변했다가 조금 변할 땐 조금만 변하는, 변동성 집중(Volatility Clustering) 이 일어나고 있습니다. 오차항이 과거의 오차항의 영향을 받고 있기 때문에 이를 식으로 표현해보겠습니다. (오차항을 과거 시점의 오차항들로 설명) $Var(\varepsilon_t) = \sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2$ 즉, 변동성(오차항의 분산)을 어제의 분산으로 설명할 수 있습니다. 이 식을 통해 오차는 다음과 같이 얻어질 수 있습니다. $\epsilon_t = w_t\sqrt{\alpha_0 + \alpha_1\varepsilon_{t-1}^2}$(w_t : 백색잡음) ← 어제의 오차가 아주 컸다면 내일의 오차도 커지겠지요? 양변에 제곱을 하면 $\epsilon_t^2 = w_t^2(\alpha_0 + \alpha_1\varepsilon_{t-1}^2)$ 가 되어 과거의 오차항을 이용해 AR type의 함수로 오늘의 오차항을 설명했습니다. 위와 같은 과정이 바로 ARCH(1) 모델입니다.

자기회귀이분산모형(ARCH)은 로버트 엥글이라는 경제학자에 의해 처음 제시된 모형으로, 과거의 정보가 주어졌을 때 현재의 값에 대한 조건부분포를 표현하기 위해 도입된 모형으로 오차의 변동성이 자귀회귀적으로 변하는 것을 설명하는 비선형 모델입니다. 다음과 같이 표현할 수 있습니다.

ARCH(p)

🥦 **ARCH(p) 수식**

$$
\begin{gathered}
Z_t \sim N(0,1) \\
\varepsilon_t = \sigma_t Z_t \\
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-1}^2 \\
= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_p\varepsilon_{t-p}^2 \\
\alpha_0 > 0,\ \alpha_j \geq 0,\ j = 1, 2, \ldots, p
\end{gathered}
$$

ARCH(p)모형은 t시점 오차항의 변동성을 p시점 전까지의 오차항의 제곱으로 설명합니다.

3. GARCH(Generalized Auto-Regressive Conditional Heteroscedasticity) : 일반자기회귀이분산모형

ARCH 모형은 과거 값의 제곱을 이용하기 때문에 방향(+/-)에 따른 영향력을 반영하지 못하며, m의 값이 커지면 추정해야 할 모수가 많아지고, 추정량의 정확도가 떨어진다는 단점이 존재합니다. 따라서 이를 해결하기 위해 제시된 일반화된 모델이 GARCH입니다. ARCH 모형이 σ_t^2에 AR 모형을 가정했던 것처럼, GARCH 모형은 σ_t^2에 ARMA 모형을 가정하여 ARCH보다 확장된 모형으로 생각하시면 됩니다!

🥦 **GARCH(p,q)**

$$
\begin{gathered}
Z_t\ N(0,1) \\
\epsilon_t = \sigma_t Z_t \\
\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 \\
= \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \cdots + \alpha_p\varepsilon_{t-p}^2 + \\
\beta_1\sigma_{t-1}^2 + \cdots + \beta_q\sigma_{t-q}^2 \\
\alpha_0 > 0,\ \alpha_j \geq 0,\ \beta_j \geq 0
\end{gathered}
$$

GARCH(p,q) 모형은 t 시점의 오차항의 변동성을 p 시점 이전의 오차항들의 제곱과 q 시점의 변동성으로 설명한 모형입니다.

Appendix

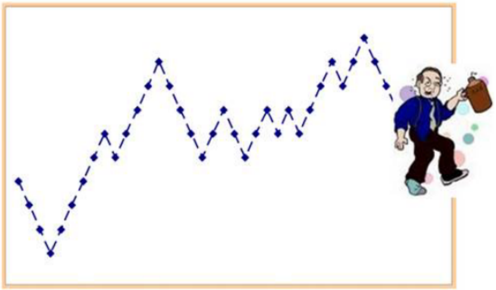
[AR 모형에서 $\phi = 1, \phi > 1$인 경우 인과성을 만족하지 못하는 이유]

*인과성: t시점의 관측값이 과거시점의 오차항으로 설명되는 특성

i. $\phi = 1$

$X_t = X_{t-1} + \epsilon_t$ or $X_t = -X_{t-1} + \epsilon_t$

이는 대표적인 비정상 확률 과정 중 하나인 확률보행과정(random walk process)입니다. $\epsilon_t \sim WN(0, \sigma^2)$일 때, $X_t = X_{t-1} + \epsilon_t$와 같이 정의되는 과정을 확률보행과정이라고 합니다. ϵ_t를 어떤 사람이 임의로 움직이는 보폭, X_t를 t시점에서의 위치라 생각한다면, X_t는 ϵ_t들의 누적합으로 표현됩니다. ($X_0 = \mu,\ X_1 = \epsilon_1, X_2 = X_1 + \epsilon_2 = \epsilon_1 + \epsilon_2,\ X_t = \sum_{i=1}^{t} \epsilon_t$) 확률보행과정은 평균은 μ로 시간에 의존하진 않지만, 자기공분산이 시간에 의존하여 비정상 확률과정입니다.



ii. $\phi > 1$

$X_t = \phi_1 X_{t-1} + Z_t = \phi_1^{M+1} X_{t-M-1} + \sum_{j=0}^{M} \phi_1^j Z_{t-j}$

위 식을 통해 알 수 있는 것처럼, $\phi > 1$일 경우 $\phi_1^{M+1} X_{t-M-1}$ 부분을 제거하지 못해 인과성을 만족하지 못합니다.